

17.9.2010

JURE-projektin sisältötyöryhmän raportti

Työryhmän kokoonpano

Jyrki Ilva, Kansalliskirjasto, pj
Aija Kaitera, HY
Marianne Kaplas, Tilastokeskus
Jukka Lindeman, THL
Sinikka Luokkanen, HAMK
Soili Makkonen, UEF
Paula Mikkonen, Kansalliskirjasto
Olli Poropudas, OKM
Tuija Raaska, CSC
Karo Salminen, Kansalliskirjasto
Tiina Salminen, TaY

1. Työryhmän työn eteneminen

JURE-projektin sisältötyöryhmä aloitti toimintansa helmikuussa 2010, ja se kokoontui syyskuun puoliväliin mennessä yhteensä viisi kertaa. Kokousten välillä työryhmän asioista käytiin keskustelua sähköpostitse.

Sisältötyöryhmälle annettiin tehtäväksi kansallisen julkaisurekisterin tietosisällön määrittely. Työryhmä lähti liikkeelle aiemmasta TiedeKOTA-työryhmän laatimasta sisältöehdotuksesta, jota muokattiin kansallisen julkaisurekisterin tarpeita varten. Taiteellista työtä koskevan tietosisällön määrittelyn todettiin vaativan vielä työryhmän ulkopuolista erityisasiantuntemusta.

Tietosisällön lisäksi työryhmä joutui pohtimaan myös järjestelmän arkkitehtuuriin ja teknisiin ratkaisuihin liittyviä kysymyksiä, jotka ovat monilta osin kytköksissä julkaisurekisterin tietosisältöön. Tältä osin työryhmällä ei ole vielä tarjottavanaan valmista esitystä, vaan sen kartoittamat vaihtoehdot toimivat pohjana jatkovalmistelulle ja päätöksenteolle.

2. Kansallisen julkaisurekisterin arkkitehtuuri ja tekniset ratkaisut

Tällä hetkellä pelkästään yliopistoissa on käytössä kymmenkunta erilaisiin teknisiin ratkaisuihin pohjautuvaa paikallista julkaisurekisteriä tai tutkimuksenhallintajärjestelmää. Osa näistä on kaupallisia tuotteita, osa taas itse kehitettyjä järjestelmiä. Julkaisutietojen tallennuspaikkana saatetaan myös käyttää esim. kirjastotietokantaa tai julkaisuarkisto-ohjelmistoa. Lisäksi joillakin organisaatioilla ei ole tällä hetkellä käytössään minkäänlaista julkaisurekisterisovellusta, vaan KOTA-raportoinnin edellyttämä julkaisutietojen keruu on hoidettu muilla keinoilla.

Kansallisen julkaisurekisterin kehittämisen kannalta teknisten sovellusten kirjo on haaste. Jotta tietojen kokoaminen nykyisistä järjestelmistä olisi mahdollista, niihin täytyisi rakentaa tiedonkeruun mahdollistavat tekniset rajapinnat ja myös järjestelmien tietosisältöä pitäisi yhdenmukaistaa. JURE:n työryhmissä onkin oltu melko yksimielisiä siitä, että kansallisen julkaisurekisterin rakentaminen on realistisinta toteuttaa vaiheittain: rekisteri perustetaan aluksi suppeassa muodossa, josta sitä voidaan sitten laajentaa.

Julkaisurekisterin arkkitehtuurin osalta työryhmä keskusteli sekä paikallisista julkaisurekistereistä

tehtävään haravointiin pohjautuvasta hajautetusta toimintamallista että kaikkien organisaatioiden yhteisen kansallisen julkaisurekisterin rakentamisesta. Lisäksi pohdittiin erilaisia mahdollisuuksia kansainvälisistä ja kotimaisista julkaisutietokannoista saatavien valmiiden viitetietojen hyödyntämiseen. Kansallisen julkaisurekisterin osalta tavoitteeksi linjattiin se, että kunkin julkaisun tiedot ovat tietokannassa vain kerran, eli eri lähdetietokannoista saadut tai eri organisaatioiden toimittamat samaa julkaisua koskevat tietueet pyritään yhdistämään kansallisen julkaisurekisterin syöttöprosessin yhteydessä toteutettavan tuplakontrollin avulla.

Hajautettu malli pohjautuisi siihen, että kukin organisaatio ylläpitäisi omaa julkaisurekisteriä, josta julkaisujen tiedot haravoitaisiin määrämuotoisina kansalliseen julkaisurekisteriin. Paikalliset järjestelmät voisivat joko olla nykyisiä tai niitä voitaisiin myös korvata jollain uudella, keskitetyllä ylläpidetyllä sovelluksella. Jälkimmäistä vaihtoehtoa perustelisi etenkin se, etteivät useimmat nykyiset järjestelmät tue valmiiden viitetietojen haravointia. Hajautetussa mallissa alkuvaiheen suurimmat haasteet liittyisivät joka tapauksessa teknisten rajapintojen rakentamiseen ja riittävän tehokkaan tuplakontrollin toteuttamiseen. Tietosisällön ja prosessien yhdenmukaistaminen olisi myös jossain määrin mutkikkaampaa, jos käytössä olisi edelleen useita erilaisia tietojärjestelmiä.

Toinen mahdollinen vaihtoehto eli yhteinen kansallinen järjestelmä puolestaan korvaisi nykyiset organisaatiokohtaiset julkaisurekisterit uudella keskitetyllä sovelluksella. Oletuksena on, että tämä uusi järjestelmä mahdollistaisi valmiiden viitetietojen hyödyntämisen. Yhteinen järjestelmä yksinkertaistaisi ja yhdenmukaistaisi muutenkin kansallisen järjestelmän prosesseja ja tuottaisi todennäköisesti merkittäviä kustannussäästöjä. Toisaalta järjestelmän yhtäaikaista käyttöönottoa koskevaan päätöksentekoon ja sen rahoituksesta sopimiseen liittyisi haasteita. Samoin järjestelmän käyttöönotto lähes yhtäaikaisesti useissa eri organisaatioissa ja sen kytkeminen moniin erilaisiin paikallisiin tietojärjestelmiin vaatisivat runsaasti koordinoitua ja käytännön työtä.

Oma kysymyksensä on kansallisen julkaisurekisterin suhde julkaisurekisteriä laajempiin organisaatiokohtaisiin tutkimuksenhallintajärjestelmiin (CRIS eli Current Research Information System). Useilla yliopistoilla on jo oma tutkimuksenhallintajärjestelmänsä jolla hallinnoidaan julkaisutietojen lisäksi myös monenlaisia muita tutkimushallinnon kannalta tarpeellisia tietoja, mikä mahdollistaa laajemman tutkimustoiminnan eri aspekteihin kohdistuvan seurannan. Näiden yliopistojen näkökulmasta siirtyminen uuteen, pelkästään julkaisutietojen hallintaan tarkoitettuun erillissovellukseen ei todennäköisesti olisi houkutteleva ratkaisu. Yhteisen kansallisen julkaisurekisterisovelluksen pitäisi siis saman tien tarjota myös muut tutkimuksenhallintajärjestelmän ominaisuudet. Tämä tarkoittaa sitä, että yhden yhteisen kansallisen järjestelmän malli edellyttäisi laajempaa keskustelua yhteisestä kansallisesta tutkimuksenhallintajärjestelmästä.

JURE-projektissa ei ole vielä tehty päätöksiä suunnitellun järjestelmän arkkitehtuurista tai teknisistä ratkaisuista. Teknisten vaihtoehtojen osalta on keskusteltu sekä CSC:n rakentamasta XDW-käsittemallia hyödyntävästä tietovarastoratkaisusta että jonkin valmiin CRIS-sovelluksen hankkimisesta kansalliseen käyttöön. XDW-pohjainen tietovarasto sopisi sinänsä hyvin haravoimalla kootun julkaisutiedon analysointiin, ja Raketti-TUTKI-hankkeessa pyritään kehittämään myös siihen pohjautuvia ratkaisuja laajempiinkin tutkimushallinnon tarpeisiin. Toisaalta tietovarasto ei sinällään sisällä esim. julkaisutietojen selailuun sopivaa www-käyttöliittymää (mitä on pidetty läpinäkyvyyden vuoksi tärkeänä), eli sen rinnalle täytyisi joka tapauksessa rakentaa jonkinlainen operatiivinen järjestelmä.

Valmiita CRIS-ohjelmistoja on markkinoilla useita, ja niiden osalta mahdollinen hankinta täytyisi toteuttaa tarjouskilpailun kautta. Keskeinen kriteeri ohjelmistovalinnassa olisi todennäköisesti se, miten hyvin se tukisi valmiiden viitetietojen hyödyntämistä. CRIS-ohjelmistoa olisi periaatteessa mahdollista hyödyntää sekä yhden yhteisen rekisterin mallissa että hajautetussa haravointiin pohjautuvassa arkkitehtuurissa. Valmisohjelmistojen kohdalla on kuitenkin syytä selvittää huolellisesti se, miten hyvin ne soveltuvat ja skaalautuvat monen organisaation yhteisen kansallisen julkaisurekisterin tarpeisiin sekä miten paljon ja millaisilla kustannuksilla niitä on mahdollista räätelöidä.

3. Valmiiden viitetietojen hyödyntäminen

Kansainvälisiä julkaisutietokantoja (Thomson-Reutersin ISI Web of Science ja Elsevierin Scopus) on tietyin ehdoin mahdollista hyödyntää myös kansallisen julkaisurekisterin lähdeaineistona ja muutenkin suomalaisen tieteellisen julkaisutoiminnan analysoinnissa. Tietokannat sisältävät valmiina huomattavan osan suomalaisten julkaisemien artikkelien viitetiedoista, ja joillakin tieteenaloilla ne ovat hyvinkin kattavia erityisesti vertaisarvioitujen kansainvälisten julkaisujen osalta. Kotimaisten julkaisujen viitetietojen osalta on perusteltua selvittää mahdollisuuksia kotimaisten viitetietokantojen (mm. Arto ja Fennica) tietosisällön ja toiminnallisuuksien kehittämiseksi siten, että ne palvelisivat nykyistä paremmin myös julkaisurekisterien tarpeita.

Kansainvälisten julkaisutietokantojen hyödyntäminen viitetietojen lähteenä kansallisen julkaisurekisterin tuotantoprosessissa on mielekästä, ja ne voivat sekä tarjota merkittäviä kustannussäästöjä tarpeettoman päällekkäisen tietojen tallentamisen vähenemisen kautta että parantaa julkaisurekisterin sisältämien tietojen laatua. Toisaalta valmiiden viitetietojen hyödyntäminen edellyttää prosessien huolellista suunnittelua ja viitetietojen haravointiin soveltuvia teknisiä ratkaisuja. Tällä hetkellä Suomessa käytössä olevista paikallisista järjestelmistä ainoastaan HY:n Tuhat tukee valmiiden viitetietojen haravointia, eli paikallisista rekistereistä ja keskitetystä rekisteristä koostuvassa hajautetussa arkkitehtuurissa haravoitujen viitetietojen hyödyntäminen edellyttäisi paikallisten järjestelmien kehittämistä tai niiden vaihtamista uuteen järjestelmään. Haasteena on myös viitetietojen ja niiden tekijöiden yhdistäminen oikeaan organisaatioon ja siellä vaikuttaviin henkilöihin. Tätä voi tiettyyn rajaan asti tehdä kansainvälisten julkaisutietokantojen viitetietoihin sisältyvien tekijöiden osoitetietojen pohjalta, mutta esim. Tuhat-järjestelmässä yhdistäminen tapahtuu lähtökohtaisesti siten, että tutkijat tunnistavat itse omat artikkelinsa.

Lisäksi on syytä pitää koko ajan mielessä, että kansainvälisten julkaisutietokantojen tiedot eivät ole kattavia, ja niiden kattavuus vaihtelee huomattavasti tieteenaloittain. Etenkin kotimaisten julkaisujen osalta kansainvälisten tietokantojen kattavuus on hyvin puutteellinen. Periaatteessa voisi ajatella, että kansainvälisistä tietokannoista saatavat viitetiedot haravoitaisiin suoraan kansalliseen tietokantaan ja tutkimusorganisaatioiden omista tietokannoista poimittaisiin mukaan ne julkaisut, joita kansainvälisistä tietokannoista ei saada. Prosessien suunnittelun kannalta tämä olisi kuitenkin haastavaa, sillä organisaatioilla pitäisi joka tapauksessa olla omassa järjestelmässään tieto myös kansainvälisistä julkaisutietokannoista haravoiduista julkaisuista, jotta ne voisivat tietää mitkä julkaisut ovat jo mukana ja mitkä vielä puuttuvat. Parhaan mahdollisen ratkaisun saavuttamiseksi tällöin olisi todennäköisesti järkevää siirtää keskitetysti haravoidut tiedot paikallisiin tietokantoihin, jossa niitä sitten täydennettäisiin. Tämä asettaisi kuitenkin lisävaatimuksia paikallisille ja keskitetyille järjestelmälle sekä niiden välisille rajapinnoille.

Samat rajoitukset koskevat myös viittaustietojen hyödyntämistä: Kansainvälisistä tietokannoista saatavat viittaustiedot koskevat vain ko. tietokantoihin sisältyviä julkaisuja, eli ne eivät anna tietoa muihin julkaisuihin mahdollisesti kohdistuvista viittauksista. Vaikka kansallista julkaisurekisteriä olisi mahdollista täydentää kansainvälisistä viitetietokannoista puuttuvien julkaisujen tiedoilla, niihin kohdistuvia viittauksia ei siis saada järjestelmään mukaan, mikä on viittausten määrien vertailemisen kannalta ongelmallista. Lisäksi kustantajien kanssa solmittavissa sopimuksissa saattaa olla rajoituksia viittaustietojen tallentamisen suhteen. Niinpä on syytä pohtia huolellisesti sitä, missä määrin viittaustietojen analysointia halutaan tehdä kansallisessa julkaisurekisterissä ja missä määrin siihen on perusteltua käyttää kansainvälisten julkaisutietokantojen omia työkaluja.

4. Kansallisen julkaisurekisterin tietosisältö

Sisältötyöryhmä kokosi esityksen kansalliseen julkaisurekisteriin sisältyvistä tiedoista (liitteet 1 ja 2). Esitys kattaa seuraavat julkaisutyyppit:

- Kausijulkaisuissa ilmestyneet artikkelit
- Monografiat
- Kirjan osat/luvut ja konferenssijulkaisuissa ilmestyneet artikkelit sekä niihin liittyvä toimitustyö

Taiteellista työtä koskevan tiedonkeruun osalta työryhmä totesi, ettei sillä ole omassa piirissään riittävää asiantuntemusta tietosisällön määrittelemiseksi.

Tietosisältöä ja kenttärakennetta määriteltäessä työryhmä oletti, että julkaisurekisterin tuotantoprosesseissa on mahdollista hyödyntää keskitetysti ylläpidettyä lehti- tai julkaisufoorumilistaa. Lista sisältäisi tietyt auktorisoidut perustiedot kaikista keskeisistä julkaisufoorumeista (tieteelliset lehdet, julkaisusarjat, mahdollisesti muitakin lehtiä ja ehkä myös kustantajia). Mahdollisuudet keskitetysti ylläpidettyjen tietojen hyödyntämiseen eri julkaisutyyppien kohdalla riippuvat jossain määrin listan kattavuudesta. Listaan liitettäviä tietoja olisivat ainakin lehden/sarjan nimi, ISSN-numero, julkaisufoorumiluokitus, julkaisumaa, lehden/sarjan sisällön open access -saatavuus sekä sen tieteenala (tai tieteenalat). Listan ja sen tietosisältöjen kokoaminen on oma erillistavoitteenaan, jonka saavuttaminen edellyttää läheistä yhteistyötä Tieteellisten Seurain Valtuuskunnan koordinoiman Julkaisufoorumi-hankkeen kanssa.

Julkaisurekisterin tietosisällön osalta työryhmä totesi edelleen olevan joitakin avoimia vaihtoehtoja, jotka kytkeytyvät laajempiin julkaisurekisterin tuotantoprosessiin tai OKM:n linjauksiin liittyviin kysymyksiin. Näillä avoimilla kysymyksillä ei ole välttämättä suurta vaikutusta tietokannan kenttärakenteeseen, mutta niillä on kuitenkin merkitystä joidenkin kenttien sisällön ja tietojen tuotantotavan kannalta.

Merkittävimmät avoimet kysymykset kytkeytyvät tieteenala- ja opintoalatietoihin, joilla on oleellinen merkitys OKM:n tarvitsemien tieteenala- ja opintoalakohtaisten raporttien tuottamisessa. Näiden tietojen osalta tarvitaan selkeitä linjauksia pyritäänkö siihen, että kukin julkaisu kytketään vain yhteen tieteenalaan (tai yhteen ensisijaiseen tieteenalaan), vai hyväksytäänkö se, että julkaisulla tai julkaisufoorumilla saattaa olla samalla kertaa useita tieteenaloja. Lisäksi tarvitaan päätöstä siitä, perustuvatko tieteenalatiedot julkaisufoorumeille määriteltäisiin tieteenaloihin, kansainvälisten julkaisutietokantojen tietoihin, organisaatioiden omaan ilmoitukseen vai jonkinlaiseen yhdistelmään näistä. Tarjolla olevat vaihtoehdot ovat myös riippuvaisia koko julkaisurekisterin tuotantoprosessista, ts. tuotetaanko se kansainvälisten julkaisutietokantojen tietojen pohjalta tai organisaatioiden omien julkaisurekisterien kautta. Sama koskee myös opintoalatietoja, olettaen etteivät ne ole suoraan konvertoitavissa tieteenalatiedoista (mitä työryhmä piti tavoiteltavana).

Julkaisutyyppien määrittelyyn liittyen työryhmä kiinnitti huomiota siihen, että kansainväliset tietokannat eivät luokittele julkaisuja samoin kriteerein kuin OKM:n julkaisutyyppiluokitus. Tietokannat eivät esim. välttämättä sisällä luotettavaa tietoa siitä, onko artikkeli käynyt läpi refereeprosessin vai ei. Tällä on merkitystä etenkin siinä tapauksessa jos kansalliseen julkaisurekisteriin haravoidaan tietoa suoraan kansainvälisistä tietokannoista.

Organisaatietietojen tuottamisen osalta keskeinen kysymys on se, perustuvatko tiedot kunkin organisaation omaan ilmoitukseen vai perustuvatko ne CSC:n tekemään kansainvälisten tietokantojen sisältämiä osoitetietoja hyödyntämään analyysiin. Ensimmäinen vaihtoehto edellyttää käytännössä sitä, että organisaatioiden ja tekijöiden tunnistaminen tapahtuu kunkin organisaation omassa järjestelmässä (tai organisaatioiden yhteisessä kansallisessa järjestelmässä). Jälkimmäisessä vaihtoehdossa taas organisaatiot voidaan tunnistaa automaattisesti niiden julkaisujen osalta joiden tiedot tulevat kansainvälisistä julkaisutietokannoista, mutta muiden julkaisujen osalta tunnistaminen täytyy tehdä muilla keinoilla.

Kansallisella tasolla tehtävää tekijöiden tunnistamista ja kansallisen tutkijarekisterin luomista työryhmä ei pitänyt tässä vaiheessa realistisena. Tekijöiden tunnistaminen ja yhdistäminen organisaatioissa toimiviin henkilöihin on mahdollista paikallisissa julkaisurekistereissä, joko niin että organisaatiossa toimivat tutkijat tunnistavat itse omat julkaisunsa (jos paikallinen

julkaisurekisterisovellus tukee tätä) tai niin että tunnistaminen tehdään virkailijatyönä. Kansallisen tietokannan tasolla tämä ei kuitenkaan ole käytännössä mahdollista ilman paikallisten julkaisurekisterien prosessien ja teknisten ratkaisujen yhdenmukaistamista tai yhteisen kansallisen julkaisurekisterin luomista. Kansainvälisistä julkaisutietokannoista poimitun datan osalta tekijätietojen yhdistäminen henkilöihin ilman tutkijoiden tai näiden kotiorganisaatioiden panosta on sen sijaan todennäköisesti liian epävarmaa ja virhealtista.

Näiden seikkojen vuoksi työryhmä piti perusteltuna sitä, että kansallinen julkaisurekisteri keskittyy ensisijaisesti keräämään tietoa siitä, mitkä organisaatiot ovat olleet mukana tuottamassa kutakin julkaisua. Niiltä osin kuin tiedot tulevat paikallisista järjestelmistä, organisaation tietojen alle voidaan tallentaa kunkin sen piirissä toimivaksi henkilöksi tunnistetun tekijän nimi. Kukin organisaatio raportoi kuitenkin tällä tavalla ainoastaan omaan henkilökuntaansa kuuluvien tekijöiden nimet. Muiden organisaatioiden tekijätiedot tulevat rekisteriin kunkin tekijän oman organisaatioiden ilmoittamina. Julkaisun täydelliset tekijätiedot tallennetaan sen sijaan bibliografisista syistä erilliseen kenttään siinä muodossa jossa ne ovat esiintyneet alkuperäisjulkaisussa tai kansainvälisissä julkaisutietokannoissa.

Julkaistujen open access -saatavuuden mittaamisen osalta ongelmaksi koettiin se, että open access voi tarkoittaa monia eri asioita. Julkaisut voivat olla saatavilla osana open access –lehteä tämän omalla verkkopalvelimella; ne voivat olla avoimesti saatavilla osana sellaista lehteä, joka muuten on pääosin maksullista aineistoa; ne saattavat olla ennen julkaisemista vapaasti saatavilla pre print -versiona jossain julkaisuarkistossa (esim. fyysikkojen ArXiv); tai ne saatetaan tallentaa julkaisemisen jälkeen avoimeen julkaisuarkistoon, jossa ne tulevat saataville joko heti tai kustantajan vaatiman embargo-ajan jälkeen. OKM:n indikaattorikäsikirjassa open access –saatavuuden määriteltiin koskevan open access –lehdessä julkaistuja artikkeleita. Tämä lienee mahdollista toteuttaa liittämällä julkaisufoorumilistaukseen tieto siitä, mitkä julkaisufoorumi ilmestyvät open access –periaatteella. Toisaalta ratkaisu ei kuitenkaan ole tyydyttävä open access –julkaisemista koskevan kokonaiskuvan luomisen kannalta, sillä se ei mittaa julkaisujen muiden reittien kautta toteutuvaa avointa saatavuutta. Työryhmä pohti sitä, pitäisikö myös tämän muun avoimen saatavuuden rekisteröinti tehdä vapaaehtoisena mahdolliseksi, vaikkei sitä vielä tässä vaiheessa edellytetäkään.

Lisäksi kansalliseen julkaisutietokantaan on perusteltua tallentaa myös lähdetietokantojen (Web of Science, Scopus, Arto, Fennica, jne.) käyttämiä tunnisteita ja ID-numeroita, joita voidaan käyttää apuvälineenä mm. tuplakontrollissa ja samaa julkaisua koskevien tietueiden yhdistämisessä.

5. Suositellut jatkotoimenpiteet

Sisältötyöryhmä totesi työnsä aikana kansallisen julkaisurekisterin rakentamisen edellyttävän huolellista suunnittelua ja riittävän väljää aikataulua, jotta optimaalisten ratkaisujen kehittäminen olisi mahdollista. Järkevänä lähestymistapana pidettiin sitä, että uusi järjestelmä rakennetaan vaiheittain siten, että rekisteri toteutetaan ensin suppeana, mahdollisuuksien mukaan jo olemassa olevia teknisiä ratkaisuja hyödyntäen, ja laajemmalle kehitystyölle määritellään selkeät pitkän tähtäimen tavoitteet ja realistinen aikataulu.

Kansallisen julkaisurekisterin arkkitehtuurin ja prosessien määrittely sekä teknisten vaihtoehtojen kartoittaminen edellyttävät vielä lisätyötä, jotta niitä koskevat päätökset ja mahdollinen hankintaprosessin käynnistäminen on mahdollista. Tarjolla on useita erilaisia, osin toisensa pois sulkevia vaihtoehtoja julkaisurekisterin arkkitehtuuriksi, joiden välillä lopulta täytyy tehdä valintoja. Paikallisiin julkaisurekistereistä haravoituihin tietoihin perustuva kansallinen julkaisurekisteri on yksi vaihtoehto, yksi kaikkien organisaatioiden yhteiseen julkaisurekisteriin tai tutkimuksenhallintajärjestelmään perustuva malli on toinen. Näiden vaihtoehtojen lisäksi arkkitehtuuria on mahdollista kehittää haravoimalla valmiita julkaisutietoja kansainvälisistä tietokannoista joko suoraan kansalliseen julkaisurekisteriin tai paikallisiin julkaisurekistereihin. Arkkitehtuuriin liittyvien ratkaisujen lisäksi tarvitaan myös tiedonkeruun ja tallennuksen prosessien

yhdenmukaistamista eri organisaatioissa, jotta tietojen yhdenmukaisuus voidaan taata.

Työryhmä piti joka tapauksessa perusteltuna sitä, että ensimmäisessä vaiheessa testataan vuoden 2010 julkaisutietojen keruuta olemassa olevista paikallisista järjestelmistä yhteiseen tietovarastoon. Tämä antaa todennäköisesti paljon tähänastista selkeämmän kuvan siitä, millaista tietoa korkeakoulujen nykyisissä järjestelmissä on, millaisessa muodossa sitä on mahdollista niistä kerätä ja miten hyvin tietojen konvertoiminen kansallisen julkaisutietokannan tarpeisiin onnistuu. Samalla olisi hyödyllistä tehdä tarkempi kartoitus korkeakouluissa julkaisurekisterikäytössä olevista tietojärjestelmistä, ja korkeakoulujen intresseistä omien julkaisurekisteriensä ja tutkimuksenhallintajärjestelmiensä kehittämisen suhteen.

Kansallisen julkaisurekisterin oheispalveluihin kuuluvan julkaisufoorumilistan (lehtilistan) suunnittelu ja kokoaminen on myös perusteltua aloittaa jo vuosien 2010-2011 aikana yhteistyössä TSV:n koordinoiman Julkaisuforumihankkeen kanssa. Listan tulee olla teknisesti yhteensopiva kansallisen julkaisurekisterin prosessien kanssa, ja siihen pitää koota julkaisurekisterin kannalta tarpeelliset tiedot.

Kansainvälisten tietokantojen osalta on syytä jatkaa selvitystyötä ja neuvotteluja siitä, miten ja millaisilla sopimusehdoilla niiden sisältämiä julkaisutietoja olisi mahdollista hyödyntää kansallisen julkaisurekisterin rakentamisessa ja OKM:n tarvitsemien julkaisutietoja koskevien analyysien tuottamisessa. Kotimaisten julkaisujen viitetietoja sisältävien kansallisten tietokantojen aineistojen osalta tarvitaan erillistä selvitystä sen suhteen, millaisia mahdollisuuksia niiden kattavuutta, tietosisältöä ja toiminnallisuuksia olisi kehittää siten, että ne palvelisivat nykyistä paremmin myös julkaisurekisterien tarpeita.

Vaikka takautuvan aineiston analyysi antaisi hyödyllistä tietoa tieteellisen toiminnan pidemmän aikavälin kehityksestä, työryhmä suhtautui kuitenkin varauksellisesti vanhojen aineistojen liittämiseen kansalliseen julkaisurekisteriin. Takautuvien aineistojen aukkojen paikkaaminen jälkikäteen on hankalaa, ja toisenlaisilla periaatteilla kerätyn tiedon konvertoiminen uuden järjestelmän tarpeisiin olisi myös haasteellista. Esim. tekijöiden tunnistaminen ja liittäminen organisaatiossa toimiviin henkilöihin on jälkikäteen tehtynä hankalaa, ja suurten aineistojen kohdalla tästä aiheutuisi runsaasti ylimääräisiä kustannuksia.

Työryhmän tuottama esitys kansallisen julkaisurekisterin tietosisällöksi sisältää vielä joitakin avoimia kysymyksiä, joiden ratkaisemiseksi tarvitaan periaatteellisia linjanvetoja ja OKM:n tiedontarpeiden tarkempaa määrittelyä. Erityisesti tämä koskee tieteenala- ja opintoalatietoja, joiden osalta merkittäviä avoimia kysymyksiä ovat mm. se, voiko sama julkaisu kuulua useampaan tieteenalaan ja millä perusteilla tiedot tuotetaan. Tehtävillä valinnoilla on merkitystä sen suhteen, miten tieteenaloja pystytään jaottelemaan järjestelmän tietojen pohjalta tuotettavissa raporteissa. Lisäksi myös taiteellista työtä koskevan tiedonkeruun osalta todettiin tarvittavien tietosisältöjen määrittämisen edellyttävän vielä erillistä selvitystyötä tai asiantuntijoista koostuvaa työryhmää.

LIITTEET:

Liite 1. Esitys Kansalliseen julkaisurekisteriin sisältyviksi tiedoiksi

Liite 2. Kansallisen julkaisurekisterin kenttien selitykset.

Liite 1: Esitys Kansalliseen julkaisurekisteriin sisältyviksi tiedoiksi (17.9.2010)

Artikkelit	Monografiat	Kirjan osat/luvut ja konferenssijulkaisut
Julkaisutyyppi	Julkaisutyyppi	Julkaisutyyppi
Tieteenala **	Tieteenala **	Tieteenala **
Opintoala **	Opintoala **	Opintoala **
Organisaatio *	Organisaatio *	Organisaatio *
Tekijä (jos henkilökuntaa) *	Tekijä (jos henkilökuntaa) *	Tekijä (jos henkilökuntaa)*
Organisaation alayksikkö *	Organisaation alayksikkö *	Organisaation alayksikkö *
Julkaisufoorumin tiedot	Julkaisufoorumin tiedot	Julkaisufoorumin tiedot
Lehden nimi ***	Julkaisija/kustantaja ***	Julkaisija/kustantaja ***
Lehden ISSN ***	Sarjan nimi ***	Sarjan nimi ***
Julkaisufoorumiluokitus ***	Sarjan ISSN ***	Sarjan ISSN ***
Julkaisumaa ***	Julkaisufoorumiluokitus ***	Julkaisufoorumiluokitus ***
Lehden avoin saatavuus ***	Julkaisumaa ***	Julkaisumaa ***
	Sarjan avoin saatavuus ***	Sarjan avoin saatavuus ***
Muut bibliografiset tiedot	Muut bibliografiset tiedot	Osan/luvun/artikkelin bibliografiset tiedot
Julkaisun nimi	Julkaisun nimi	Julkaisun nimi
Julkaisun täydelliset tekijätiedot	Julkaisun täydelliset tekijätiedot	Julkaisun täydelliset tekijätiedot
Julkaisuvuosi	Julkaisun ISBN	Sivut
Volyymi	Julkaisuvuosi	
Numero	Sivuja yhteensä	
Sivut		Emojulkaisun bibliografiset tiedot
		Emojulkaisun nimi
Täydentävät tiedot	Täydentävät tiedot	Emojulkaisun toimittajat
Julkaisun kieli	Julkaisun kieli	Emojulkaisun ISBN
Julkaisun pysyvä verkko-osoite *	Muu avoin saatavuus *	Julkaisuvuosi
Muu avoin saatavuus *	Julkaisun pysyvä verkko-osoite*	
Lähdetietokannan tunniste *	Lähdetietokannan tunniste *	Täydentävät tiedot
		Julkaisun kieli
		Julkaisun pysyvä verkko-osoite *
		Muu avoin saatavuus *
		Lähdetietokannan tunniste *

* Toistettavia kenttiä

** Toistetaanko näitä myös (ts. voiko sama julkaisu kuulua useampaan tieteen/opintoalaan?)?

*** Julkaisufoorumilistaan sisältyviä tietoja

Liite 2. Kansallisen julkaisurekisterin kenttien selitykset

Artikkelit

Julkaisutyyppi

Perustuu OKM:n KOTA-tiedonkeruussa käytettyyn julkaisutyyppiluokitukseen.

Tieteenala

Toistettava kenttä

Yhteensopiva Tilastokeskuksen tieteenalaluokituksen ja Julkaisuforumihankkeessa käytetyn luokituksen kanssa. Tieto perustuu primäärisesti lehdelle julkaisuforumilistalla määritellyyn tieteenalaan, toissijaisesti (jos tieto ei ole saatavissa em. lähteestä) kansainvälisen julkaisutietokannan luokitukseen tai organisaation ilmoitukseen. *Ei-tieteellisten julkaisutyyppien osalta tieto voi olla myös vapaaehtoinen.*

Opintoala

Toistettava kenttä

Yhteensopiva Opetushallinnon koulutusalaaluokituksen kanssa. Pyritään siihen, että tieto olisi mahdollista konvertoida tieteenalatiedon pohjalta, jotta sen tuottamiseen ei tarvita erillistä prosessia.

Organisaatio

Toistettava kenttä

Suomessa toimiva tutkimusorganisaatio (yliopisto, ammattikorkeakoulu, tutkimuslaitos), jonka tutkija tai tutkijoita on osallistunut julkaisun tekemiseen. Jos tiedot kootaan hajautetusti eri organisaatioista, kukin organisaatio ilmoittaa vain oman organisaationsa tiedot.

Tekijä

Organisaation alakenttä

Toistettava kenttä

Organisaatioon kuuluvat tutkijat, jotka ovat osallistuneet julkaisun tekemiseen. Jos mahdollista, tekijätiedoissa esiintyvä henkilönimi on tunnistettu ja yhdistetty organisaatiossa toimivaan henkilöön. Ilmoitetaan muodossa Sukunimi, Etunimi (tai Etunimet).

Organisaation alayksikkö

Organisaation alakenttä

Toistettava kenttä

Vapaaehtoinen kenttä

Organisaatiot voivat halutessaan ilmoittaa myös niiden oman organisaationsa sisäisten yksiköiden (esim. osasto, tdk, laitos) nimet, joissa tekijät toimivat. Yhden alayksikön tiedot tallennetaan samaan kenttään eikä niitä yhdistetä tekijätietoihin.

Lehden nimi

Yhteys julkaisuforumilistaan

Julkaisuforumina käytetyn lehden nimi. Ilmoitetut tiedot tarkistetaan keskitetysti ylläpidetystä julkaisuforumilistasta.

Lehden ISSN

Yhteys julkaisuforumilistaan

Julkaisuforumina käytetyn lehden ISSN-numero. Ilmoitetut tiedot tarkistetaan keskitetysti ylläpidetystä julkaisuforumilistasta.

Julkaisufoorumiluokitus

Tulee julkaisufoorumilistasta

Julkaisufoorumin luokitus. Tieto poimitaan keskitetysti ylläpidetystä julkaisufoorumilistasta.

Julkaisumaa

Yhteys julkaisufoorumilistaan

Lehden julkaisumaa. Ilmoitetut tiedot tarkistetaan keskitetysti ylläpidetystä julkaisufoorumilistasta.

Lehden avoin saatavuus

Tulee julkaisufoorumilistasta

Toimiiko lehti open access –periaatteella? Tieto poimitaan keskitetysti ylläpidetystä julkaisufoorumilistasta.

Julkaisun muu avoin saatavuus

Vapaaehtoinen kenttä

Toistettava kenttä

Riippumatta siitä onko julkaisufoorumi (=lehti) sinällään avoin tai ei, tieto muuta kautta toteutuvasta avoimesta saatavuudesta voidaan ilmoittaa tässä kentässä. *Rinnakkaistallennuksen embargo-aikojen, pre printtien ym. osalta tarvitaan vielä linjauksia. Onko yhdistettävissä edelliseen kenttään?*

Julkaisun pysyvä verkko-osoite

Toistettava kenttä

Julkaisun pysyviin tunnisteisiin (esim. DOI, URN tai handle) pohjautuva verkko-osoite, joka vie suoraan julkaisun kokotekstiversioon (ei siis esim. lehden/kustantajan etusivulle). Voi viitata sekä vapaasti saatavilla olevaan versioon että kustantajan palvelimella sijaitsevaan käyttöoikeudeltaan rajoitettuun versioon.

Julkaisun nimi

Julkaisun (=artikkelin) nimi. Pääotsikko ja mahdollinen alaotsikko tallennetaan samaan kenttään, ja ne voi erottaa toisistaan välilyönti, kaksoispiste, välilyönti -yhdistelmällä. Vieraskielisen julkaisun nimi voidaan tarvittaessa tallentaa translitteroituna versiona.

Täydelliset tekijätiedot

Julkaisun täydelliset tekijätiedot (ml. ulkomaiset tekijät) siinä muodossa ja järjestyksessä jossa ne on listattu alkuperäisessä julkaisussa tai lähdetietokannoissa.

Julkaisuvuosi

Julkaisun ilmestymisvuosi. Ilmoitetaan lehden vuosikerran mukaan.

Volyyymi

Lehden volyyymi, jossa artikkeli on ilmestynyt.

Numero

Lehden numero, jossa artikkeli on ilmestynyt.

Sivut

Numeron tai volyymin sivunumerot (esim. 1-20), joilla artikkeli on ilmestynyt.

Julkaisun kieli

Kieli, jolla artikkeli on kirjoitettu. Mahdollisen muunkielisen tiivistelmän kieltä ei mainita.

Lähdetietokannan tunniste

Toistettava kenttä

Julkaisun tunniste tai ID-numero tietokannassa, josta sen tietue on haravoitu.

Monografiat

Pääosa kentistä on yhteneväisiä kohdan Artikkelit kanssa.

Julkaisutyyppi

ks. Artikkelit

Tieteenala

ks. Artikkelit

Niiltä osin kuin julkaisut eivät ole ilmestyneet jonkin julkaisufoorumilistaan sisältyvän sarjan osana tai julkaisufoorumilistaan sisältyvän kustantajan julkaisemana, niiden tieteenala täytyy määritellä tapauskohtaisesti.

Opintoala

ks. Artikkelit

Organisaatio

ks. Artikkelit

Tekijä

ks. Artikkelit

Organisaation alayksikkö

ks. Artikkelit

Sarjan nimi

Yhteys julkaisufoorumilistaan

Sarja, jossa monografia on julkaistu. Ilmoitetut tiedot tarkistetaan keskitetysti ylläpidetystä julkaisufoorumilistasta.

Sarjan ISSN

Yhteys julkaisufoorumilistaan

Monografian julkaiseen sarjan ISSN. Ilmoitetut tiedot tarkistetaan keskitetysti ylläpidetystä julkaisufoorumilistasta.

Julkaisija/Kustantaja

Yhteys julkaisufoorumilistaan

Julkaisun kustantaja. Ilmoitetut tiedot tarkistetaan keskitetysti ylläpidetystä julkaisufoorumilistasta.

Julkaisufoorumiluokitus

Tulee julkaisufoorumilistasta

Lehden tai kustantajan julkaisufoorumiluokitus. Tieto poimitaan keskitetysti ylläpidetystä julkaisufoorumilistasta.

Julkaisumaa

Yhteys julkaisufoorumilistaan

Maa, jossa julkaisu on ilmestynyt. Ilmoitetut tiedot tarkistetaan keskitetysti ylläpidetystä julkaisufoorumilistasta.

Sarjan avoin saatavuus

Tulee julkaisufoorumilistasta

Toimiiko sarja open access –periaatteella? Tieto poimitaan julkaisufoorumilistasta.

Muu avoin saatavuus

ks. Artikkelit

Julkaisun pysyvä verkko-osoite

ks. Artikkelit

Julkaisun nimi

ks. Artikkelit

Julkaisun ISBN

Julkaisun (monografian) ISBN-numero.

Julkaisun numero

Julkaisun järjestysnumero sen julkaisseessa sarjassa.

Täydelliset tekijätiedot

ks. Artikkelit

Julkaisuvuosi

Julkaisun ilmestymisvuosi.

Sivuja yhteensä

Julkaisun sivumäärä (esim. 300).

Julkaisun kieli

ks. Artikkelit

Lähdetietokannan tunniste

ks. Artikkelit

Kirjan osat/luvut ja konferenssijulkaisut

Pääosa kentistä on yhteneväisiä kohdan Artikkelit kanssa.

Julkaisutyyppi

ks. Artikkelit

Tieteenala

ks. Artikkelit

Niiltä osin kuin julkaisut eivät ole ilmestyneet jonkin julkaisufoorumilistaan sisältyvän sarjan osana tai julkaisufoorumilistaan sisältyvän kustantajan julkaisemana, niiden tieteenala täytyy määritellä tapauskohtaisesti.

Opintoala

ks. Artikkelit

Organisaatio

ks. Artikkelit

Tekijä

Organisaation alakenttä

Toistettava kenttä

Organisaatioon kuuluvat tutkijat, jotka ovat osallistuneet julkaisun tekemiseen. Jos mahdollista,

tekijätiedoissa esiintyvä henkilönimi on tunnistettu ja yhdistetty organisaatiossa toimivaan henkilöön. Ilmoitetaan muodossa Sukunimi, Etunimi (tai Etunimet). Kukin osajulkaisu raportoidaan erikseen, vaikka emojulkaisu olisi sama. Samoin toimitustyö raportoidaan erikseen. Tekijän rooli toimittajana ilmenee tällöin Julkaisutyyppi-kentässä.

Organisaation alayksikkö

ks. Artikkelit

Sarjan nimi

Yhteys julkaisuforumilistaan

Sarja, jossa emojulkaisu on ilmestynyt. Ilmoitetut tiedot tarkistetaan keskitetysti ylläpidetystä julkaisuforumilistasta.

Sarjan ISSN

Yhteys julkaisuforumilistaan

Emojulkaisun julkaisseen sarjan ISSN. Ilmoitetut tiedot tarkistetaan keskitetysti ylläpidetystä julkaisuforumilistasta.

Julkaisija/Kustantaja

Yhteys julkaisuforumilistaan

Sarjan/emojulkaisun kustantaja. Sarjan osalta ilmoitetut tiedot tarkistetaan keskitetysti ylläpidetystä julkaisuforumilistasta.

Julkaisuforumiluokitus

Tulee julkaisuforumilistasta

Julkaisuforumin luokitus. Tieto poimitaan julkaisuforumilistasta.

Julkaisumaa

Yhteys julkaisuforumilistaan

Maa, jossa sarja/emojulkaisu on ilmestynyt. Sarjan osalta ilmoitetut tiedot tarkistetaan keskitetysti ylläpidetystä julkaisuforumilistasta.

Sarjan avoin saatavuus

Tulee julkaisuforumilistasta

Toimiiko sarja open access –periaatteella? Tieto poimitaan julkaisuforumilistasta.

Muu avoin saatavuus

ks. Artikkelit

Julkaisun pysyvä verkko-osoite

ks. Artikkelit

Linkki voi viedä myös emojulkaisuun, jos linkittäminen osajulkaisuun ei ole mahdollista.

Emojulkaisun nimi

Emojulkaisun (esim. artikkelikokoelma) nimi.

Emojulkaisun toimittajat

Emojulkaisun toimittajien nimet siinä muodossa ja järjestyksessä jossa ne on listattu alkuperäisessä julkaisussa tai lähdetietokannassa.

Emojulkaisun ISBN

Emojulkaisun ISBN-numero.

Julkaisun nimi

ks. Artikkelit

Täydelliset tekijätiedot

Osan/luvun täydelliset tekijätiedot (ml. ulkomaiset tekijät) siinä muodossa ja järjestyksessä jossa ne on listattu alkuperäisessä julkaisussa tai lähdetietokannassa.

Sivut

Osan/luvun sivunumerot emojulkaisussa (esim. 1-20).

Julkaisu vuosi

Emojulkaisun ilmestymisvuosi.

Julkaisun kieli

ks. Artikkelit.

Lähdetietokannan tunniste

ks. Artikkelit